CmpSpots - 2D Spot Pairing
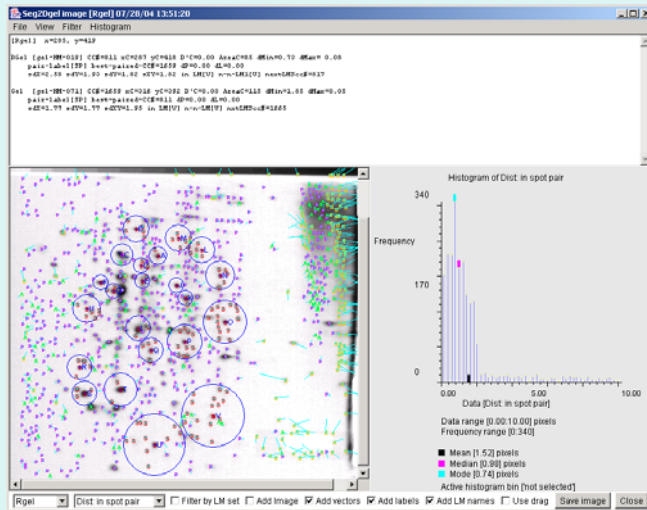
Outline

1. Overview
2. Algorithm
   - 2.1 Landmark sets
   - 2.2 Pairing features dP and dL
   - 2.3 Primary spot pairing
   - 2.4 Secondary pairing
3. Web page
4. Report window GUI
5. Options Wizard GUI
6. Image Viewer - GUI
7. Image Viewer - Examples

Overview

- CmpSpots is an open-source Java 2D quantified spot pairing program for matching spots between two lists of spots from 2D data samples.

- "Spot" data could be from 2D electrophoretic gel images, 2D LC-MS peptide clusters, or compatible data source, etc.

- There may be different numbers of spots in the two samples implying missing data or N:1 or 1:N spot matching.

- CmpSpots is a step [4] module in the pipeline analysis for the Open2Dprot project.

- Spot data may be used from the Open2Dprot Seg2Dgel 2D PAGE gel spot quantification program or compatible source.

Overview - Continued (2)

- Input spot lists are read from Sample Spot-list Files (SSF) XML (or tab-delimited) files and paired-spot lists are generated in XML, tab-delimited, or human readable formats as Sample Comparison Files (SCF).

- Uses a sample description experiment "accession" database listing the gels to be used and their region of interests. This DB will be replaced with a MIAPE compliant XML database format and a tool added to Open2Dprot.

- Uses a set of predefined landmarks between the sample gel and a reference gel from a landmark database. This DB will be interactively defined as an XML DB in the future using a new Open2Dprot tool.

4

Overview - Continued (3)

- The program may be run either interactively (-gui) with a graphical user interface (GUI) or under an OS shell command to implement batch (-nogui).

- Run-time options are specified as Unix-style command line '-' prefixed input switches so the program is also usable under batch. These may be set by an Options Wizard window.

- In the GUI version, after the spot pairing is finished, the user has the option of interactively viewing the paired spot data generated by the spot pairing with the Image Viewer window.

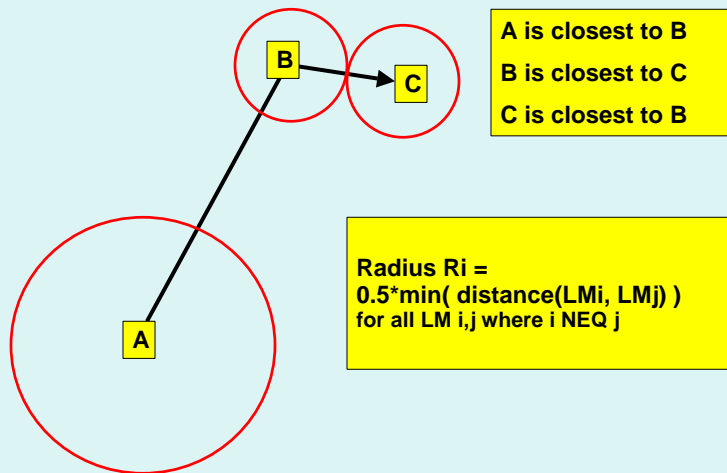- Web site contains documentation, downloads and examples.

5

Algorithm (1)

1. Three parameters are required: thrSP and thrPP the maximum distance between warped putative spots for them to be labelled as Sure-Pairs and as Possible-Pairs; number of alternative landmarks to check when doing secondary pairing.

2. Two input spot lists are read from XML files for a Reference sample and the sample being matched.

3. The landmarks for these samples area read from a landmark database and landmarks effective radii computed.

$$R_k = 1/2 \, Min(\, distance(Lm_k, Lm_m) \,)$$
for all landmarks k,m where k not m

4. Spots are assigned to closest landmark as landmark sets, as well as noting their next-nearest landmarks.

6

Effective Radii of Landmarks

Estimates of spot pairs within landmark radii are more reliable

B

C

A

A is closest to B

B is closest to C

C is closest to B

Radius Ri =
0.5*min(distance(LMi, LMj))
for all LM i,j where i NEQ j

Definition of dP - distance between spots in a pair

The <u>distance between paired spots</u> is **dP_{kij}** and is computed after the two spots (s_{1i}, s_{2j}) are mapped to (s'_{1i}, s'_{2j}) in the same sample space relative to landmark k.

$$dP_{kij} = |s'_{1i} - s'_{2j}|.$$

s'_{1i}

dP_{kij}

s'_{2j}

dL_{kij}

+ LMk
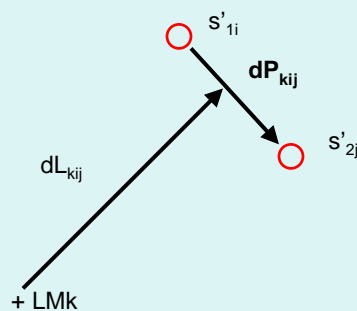
Definition of dL - distance from a pair to it's landmark

The <u>distance between a spot-pair and the landmark</u> is **dL$_{kij}$** and is computed after the two spots (s_{1l},s_{2j}) are mapped to (s'_{1l},s'_{2j}) in the same sample space relative to landmark k.

dL_{k1i} = distance(LMk,s'_{1i}),
dL_{k2j} = distance(LMk,s'_{2j}),
dL_{kij} = min(dL_{k1i}, dL_{k2j}).

s'_{1i}

dP_{kij}

dL$_{k1j}$

dL$_{kij}$

s'_{2j}

dL$_{k2j}$

+ LMk
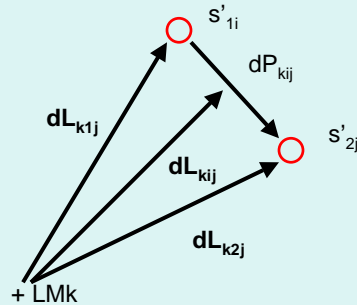
9

Algorithm (continued) (2)

5. For each landmark, spots are optimally paired between the corresponding landmark sets:

   a) mutually-paired spots I,j assigned sure-pair (SP) labels if (dP_{kij} <= thrSP) and (dL_{kij} <= R_k).

   b) mutually-paired spots assigned possible-pair (PP) labels if (dP_{kij} <= thrPP) or ((dP_{kij} <= thrSP) and (dL_{kij} > R_k)).

   c) N:1 or 1:N paired spots assigned ambiguous-pair (AP) labels if N spots in one sample pair with 1 spot in other sample.

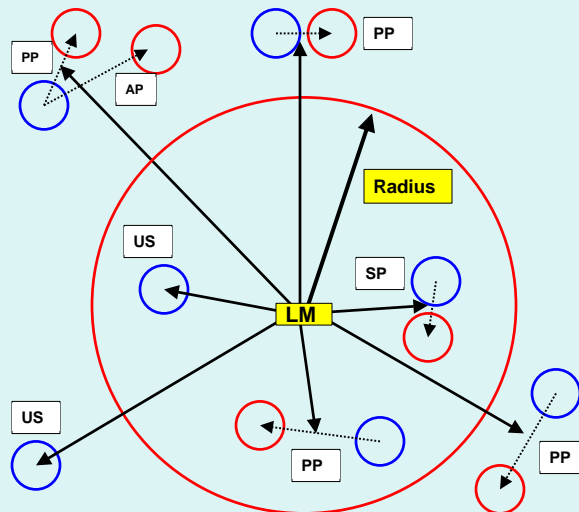   d) Unresolved Spots (US) are what remains unpaired.

10

Primary Pairing Cases

Case
[1] US is unresolved spot (no dP),
[2] SP is dL <= R_k and dP <= dTsp,
[3] PP is dL >= R_k and dP <= dTpp,
[4] PP is dL <= R_k and dP >= dTsp and dP <= dTpp,
[5] PP is dL >= R_k and dP <= dTsp,
[6] PP is dL >= R_k and dP <= dTpp. For the other spot
   AP' is dL' >= R_k and dP' <= dTpp and dP' >= dP

Primary Pairing Cases

Algorithm (continued) (3)

6. Secondary spot pairing can be used to further resolve AP and US labels in adjacent landmark sets into SP or PP labels which are then placed in either of the two sets. There are four cases: (a) two unresolved spots (US and US), (b) two ambiguous pairs (AP and AP), (c-d) one ambiguous spot (AP) and one unresolved spot (US).

   If the new spot pair is better than its previous pairing, it is upgraded to a SP or PP pairing label and is put into whichever landmark set has the smallest dL for the putative pair.

7. The data is then saved in a SPF output file for further processing   by other programs such as the Open2Dprot database builder.

13

Secondary Pairing Cases

- A second pass is optimizes the spot pairing to further resolve AP and US labels in adjacent landmark sets into PP labels that are then placed in either of the two sets.

- There are four cases (see Reference Manual for details):
    (a) two unresolved spots (US and US),
    (b) two ambiguous pairs (AP and AP),
    (c-d) one ambiguous spot (AP) and one unresolved spot (US).

- The new spot pair is upgraded to a SP or PP pairing label and is put into whichever landmark set has the smallest dL for the putative pair.

- Each landmark set is tested. Each spot has a list of adjacent landmarks, so these are tested to see if pairing can be optimized for that spot.

14

Secondary Pairing Cases

15
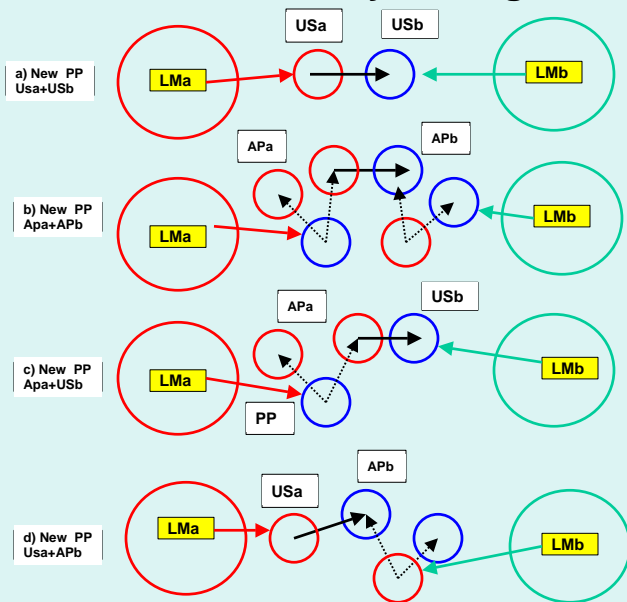
In Table of Contents:

* **Introduction**

* **Overview (PDF)**

* **Reference Manual**

* **Demos & Screen shots**

* **Download**
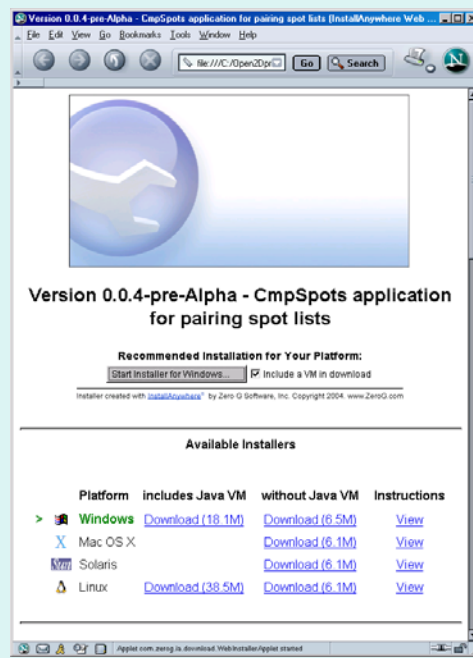
* **Source code**

16

Downloading and Installing CmpSpots

- You may <u>download</u> the pre-alpha version and install it on your computer from the Web site.

- Currently, CmpSpots is hardwired to start with the demo gel and with the -gui switch. However, you can override this with the **Edit options** popup wizard.

17

Installing CmpSpots on your computer

18

CmpSpots Report Window Interface in -gui mode

Pull-down menus

Report window text

Command buttons

19

CmpSpots Report Window Interface in -gui mode
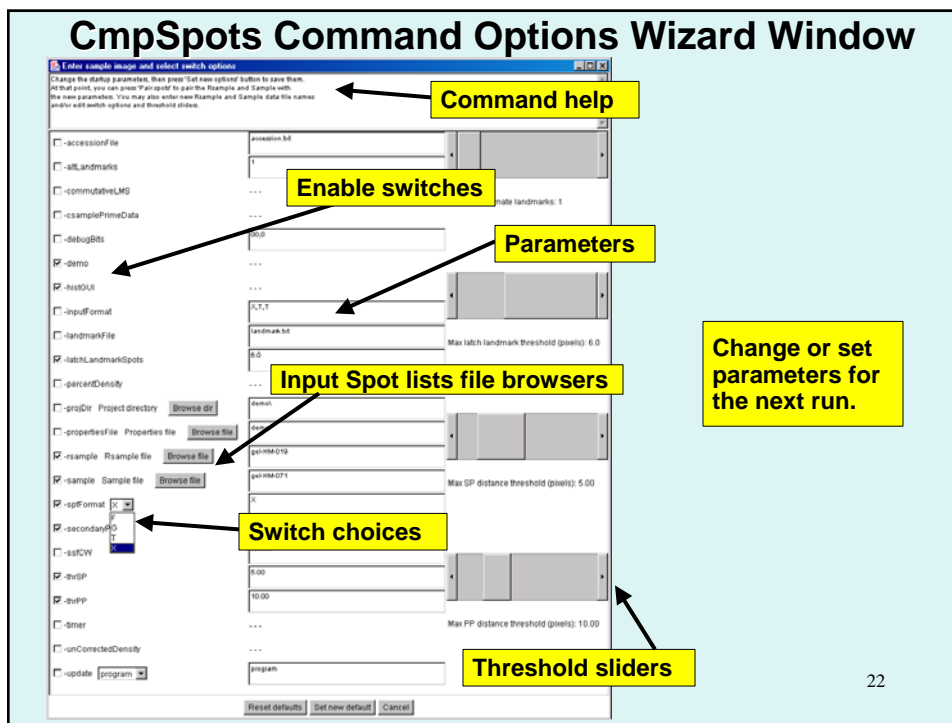

Pull-down menus

Report window text

Command buttons

20

Option Wizard to Edit Command Options

- The user may modify the command line switch options using the Option Wizard by pressing the **Edit Options** button.

- You may save the new options in a "*CmpSpots.properties*" file in the current project directory .

- The new options are then used as the new default switch options in subsequent running of CmpSpots.

21

CmpSpots Command Options Wizard Window

Command help

Enable switches

Parameters

Change or set parameters for the next run.

Input Spot lists file browsers

Switch choices

Threshold sliders

22

Image Viewer Window

- Interactively view the spot pairing results overlayed on the original images using the Image Viewer.

- Select spot to examine the features of the spot-pair.

- View (landmarks, pairing-labels, spot-pairing vectors, landmark radii) as overlays.

- View subsets of above for currently selected landmark.

- Dynamic histograms spot-pairing statistics (dP, dL, pairing-labels, LM set size) features.

- Filter by pairing-labels and/or histogram features.

23

Data used in the following examples

Rsample (G1=Leukemia-AML*) has 933 in all landmark sets,
Sample (G2=Leukemia-ALL*) 2143 spots in all landmark sets.

After <u>Initial pairing:</u>
US 853
SP 242
PP 1028
AP 792
CP 0
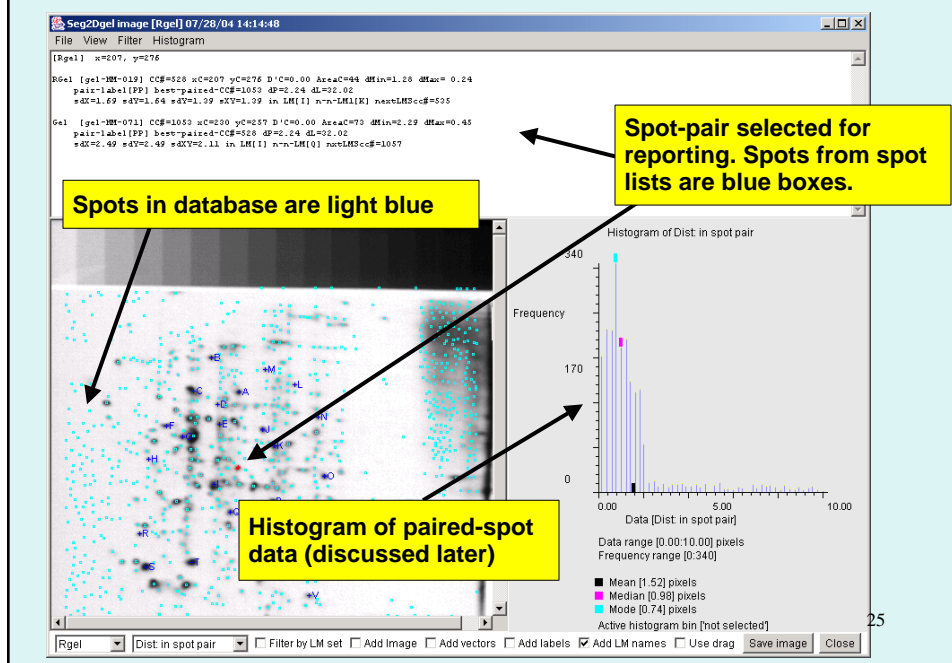EP 0
0.5(SP+PP)/(|G1| MIN |G2|)=68.06%

After <u>secondary pairing:</u>
US 834
SP 242
PP 1082
AP 757
CP 0
EP 0

After secondary pairing: 0.5(SP+PP)/(|G1| MIN |G2|)=71.0%
mean dP(SP+PP)=4.66, mean dP'((|G1|+|G2|)/(SP+PP))=8.39

*Data from CmpSpots demo database

24

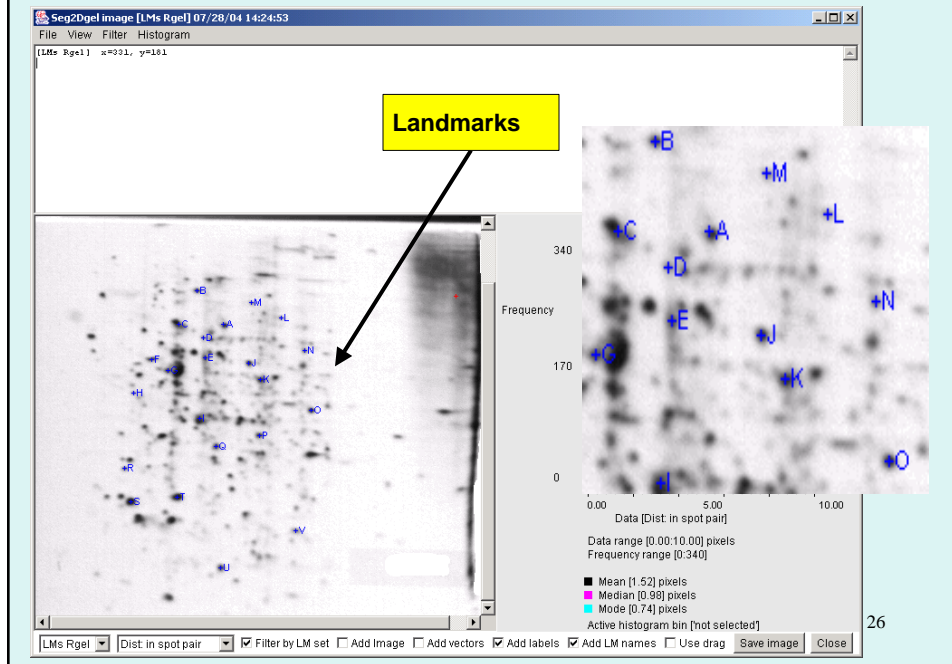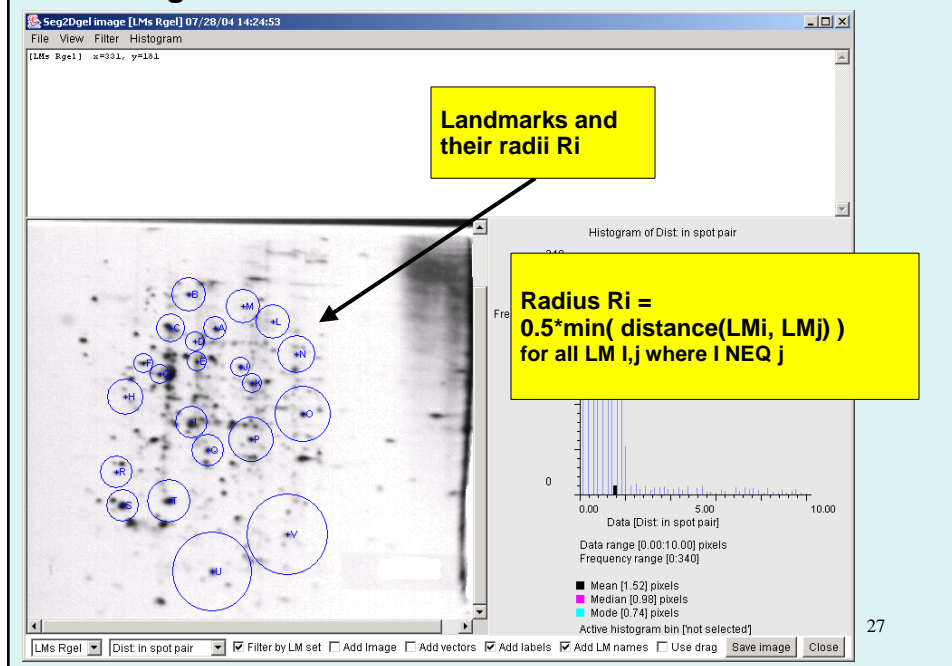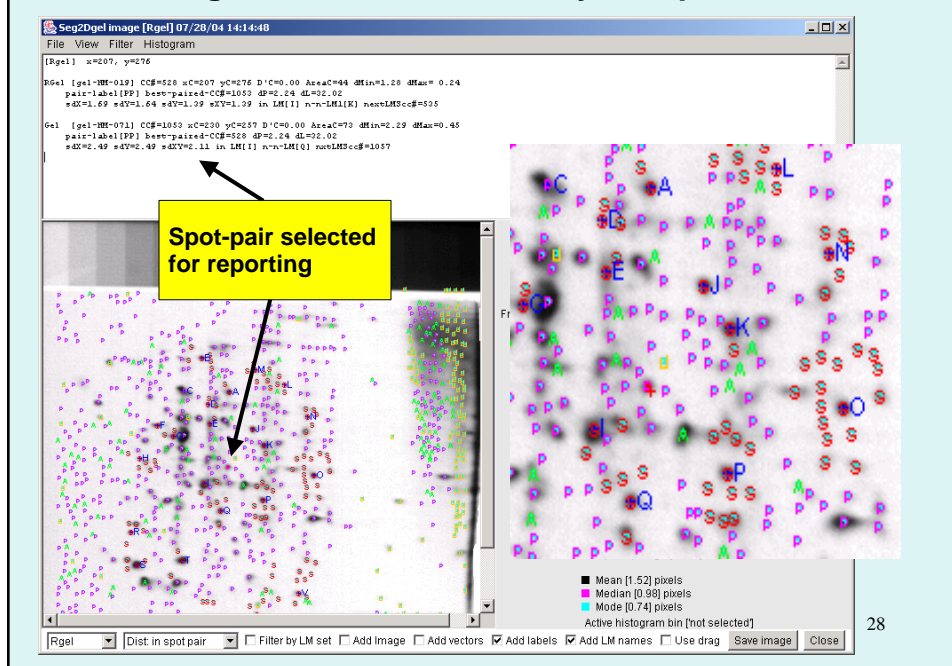**Image Viewer - no overlays, no spot filters**

Spot-pair selected for reporting. Spots from spot lists are blue boxes.

Spots in database are light blue

Histogram of paired-spot data (discussed later)

Image Viewer - landmarks

Landmarks
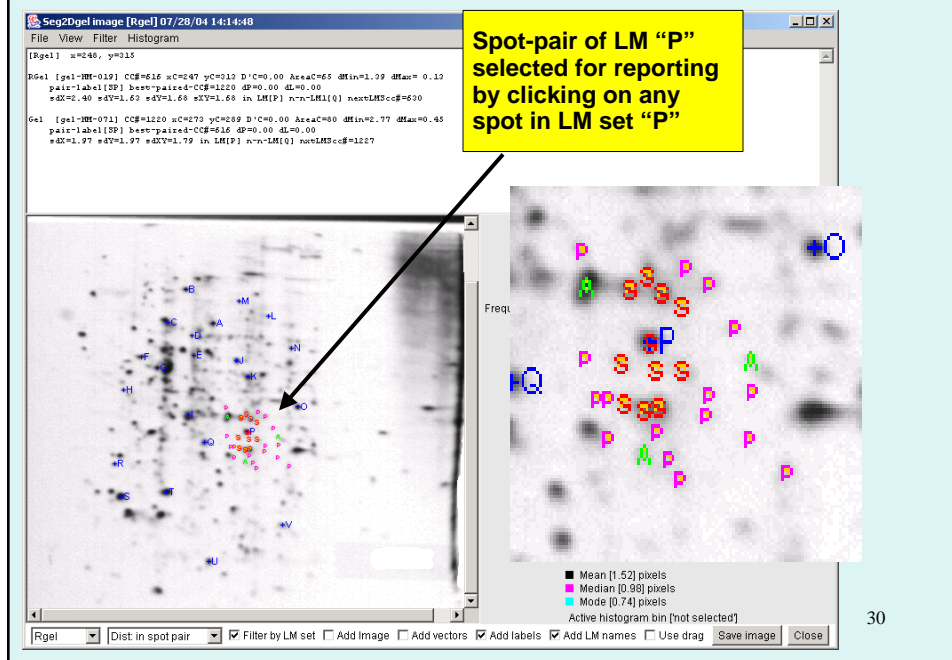
Image Viewer - landmarks and their effective radii

Landmarks and their radii Ri

Radius Ri = 0.5*min(distance(LMi, LMj)) for all LM I,j where I NEQ j

27

Image Viewer - labels overlay, no spot filters

Spot-pair selected for reporting

28

Image Viewer - labels and landmarks radii overlays

Labelled spot pairs and landmarks radii

29
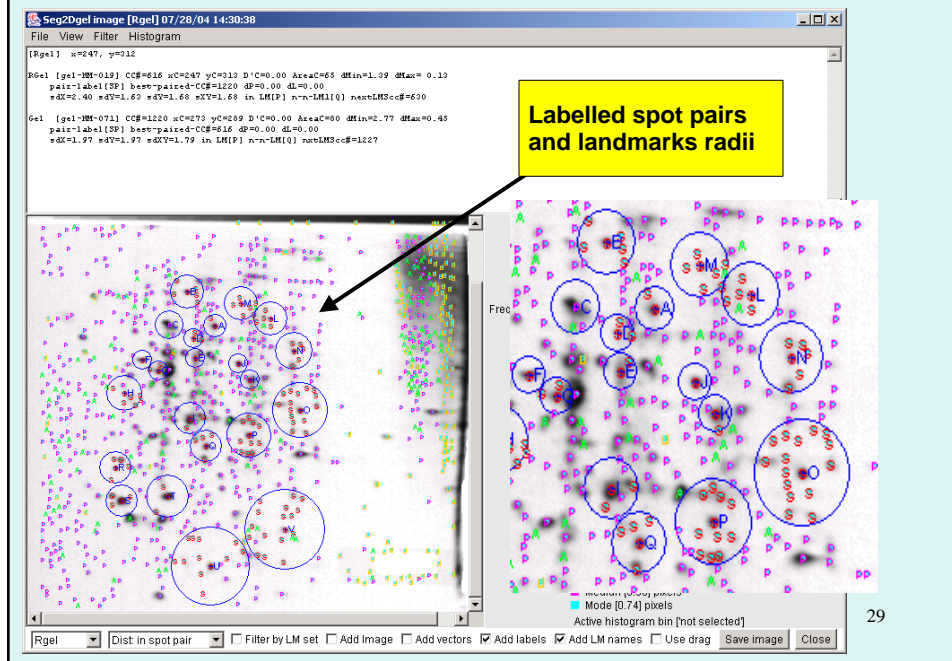
Image Viewer - labels overlay, current landmark set "P"

Spot-pair of LM "P" selected for reporting by clicking on any spot in LM set "P"

30

Image Viewer - labels & radius overlay, landmark set "P"

Spot-pair of LM "P" selected for reporting with LM radius
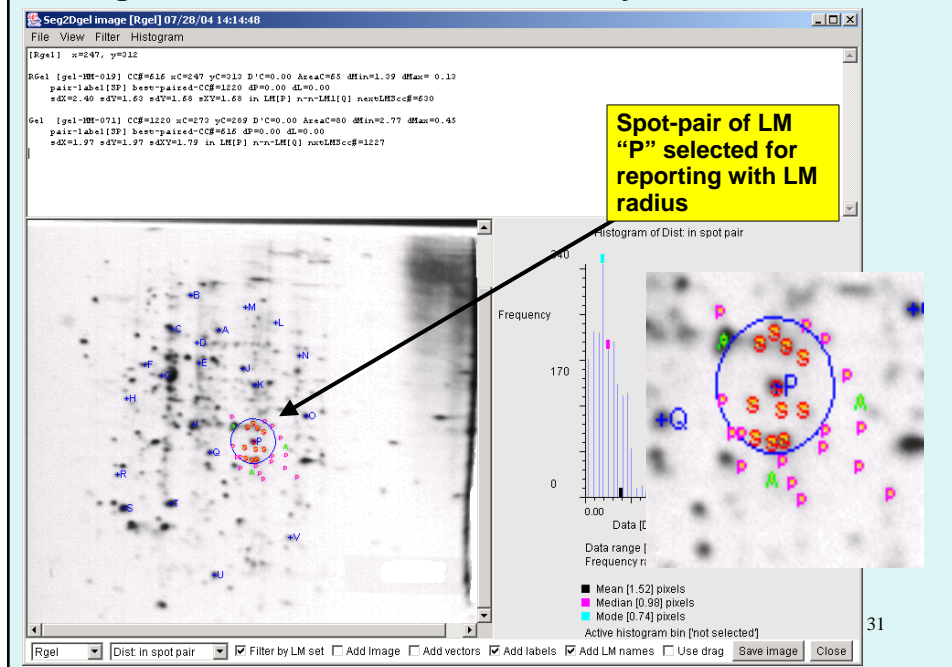
31

Image Viewer - vector difference overlay, no filter

Poor pairing has longer vectors
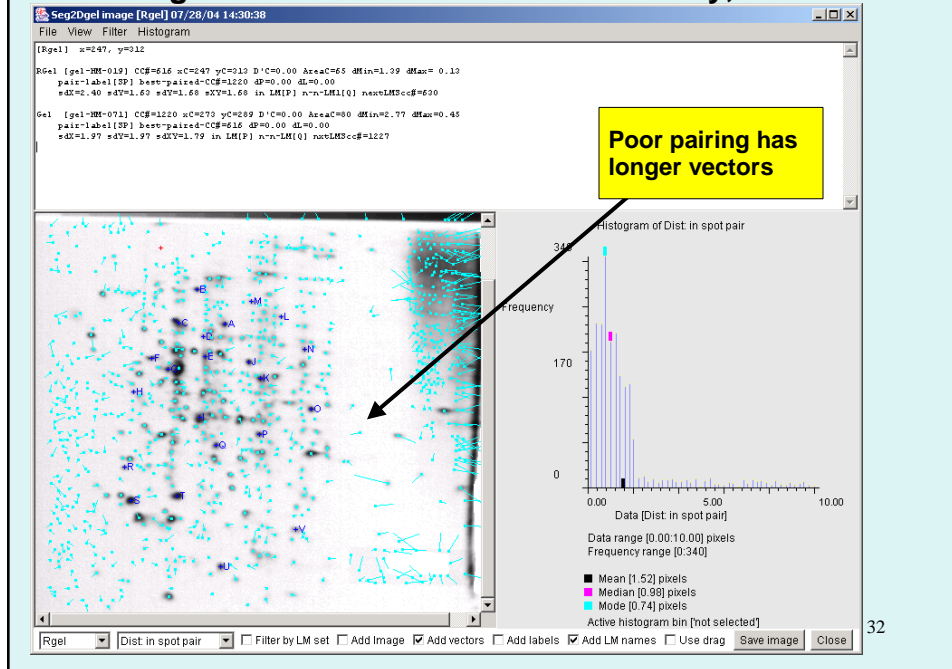
32

Image Viewer - vector overlay, landmark set "V"

Pairs further away from landmarks have worse pairing - mostly US spots

Image Viewer - labels &radius overlay, landmark set "V"

Pairs closer to landmarks have better pairing - mostly SP or PP
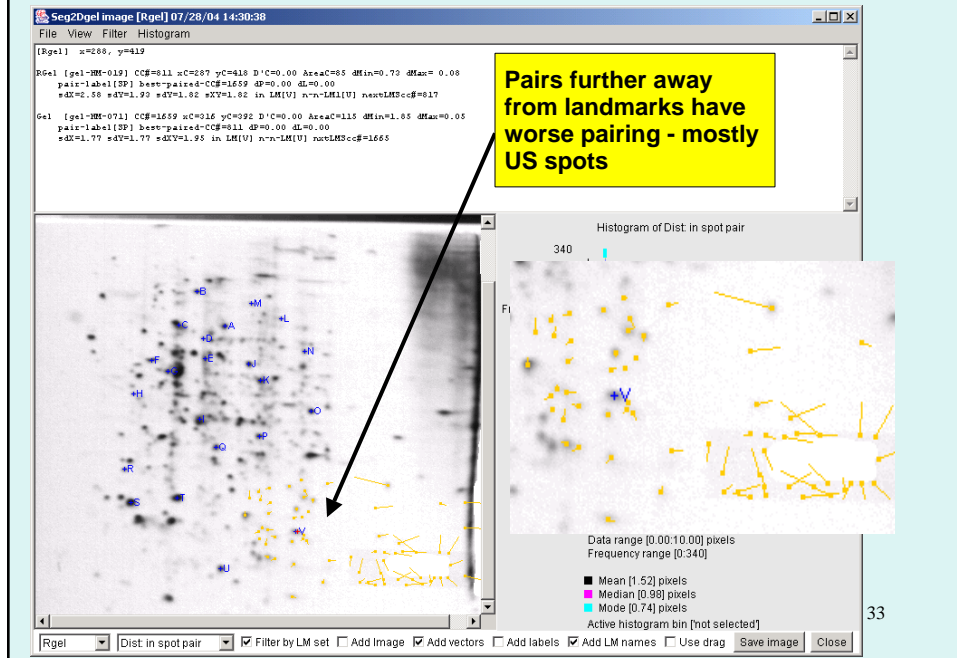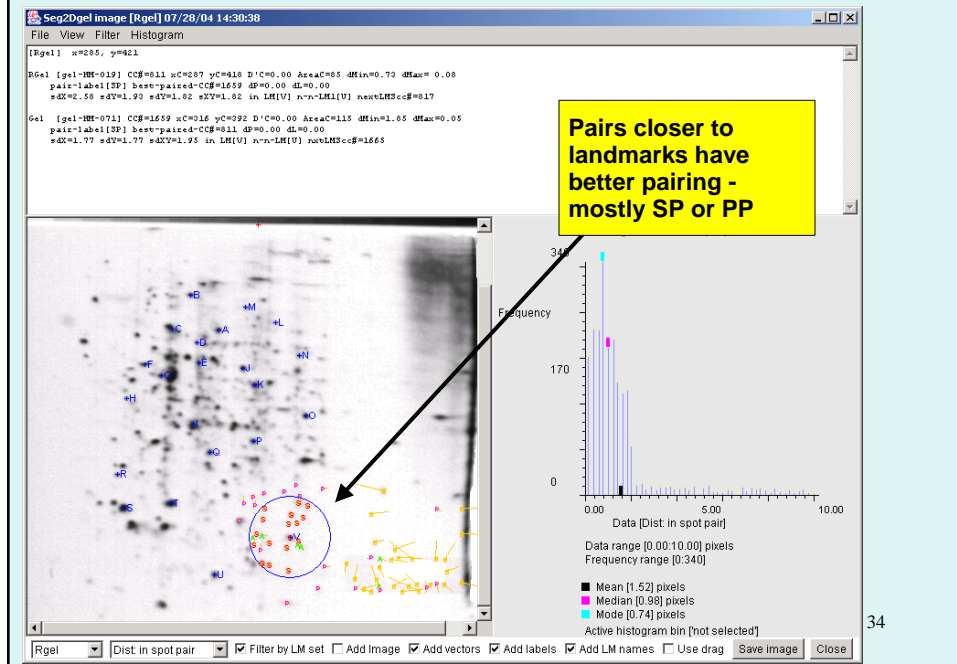
Image Viewer - landmark sets "A, B, C, D"
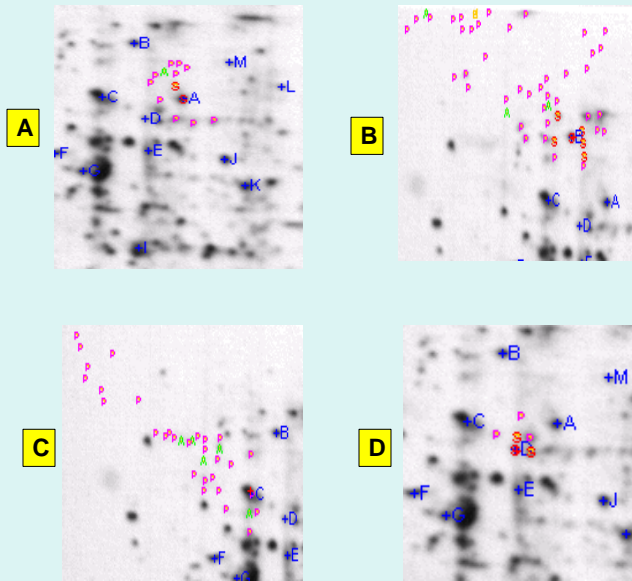
35

Image Viewer - histogram of pairing features

1. Filter spot data by spot-pair feature histogram bins

3. Filtered spot-pairs shown in blue

2. Select bin to see value or filter spot-pairs
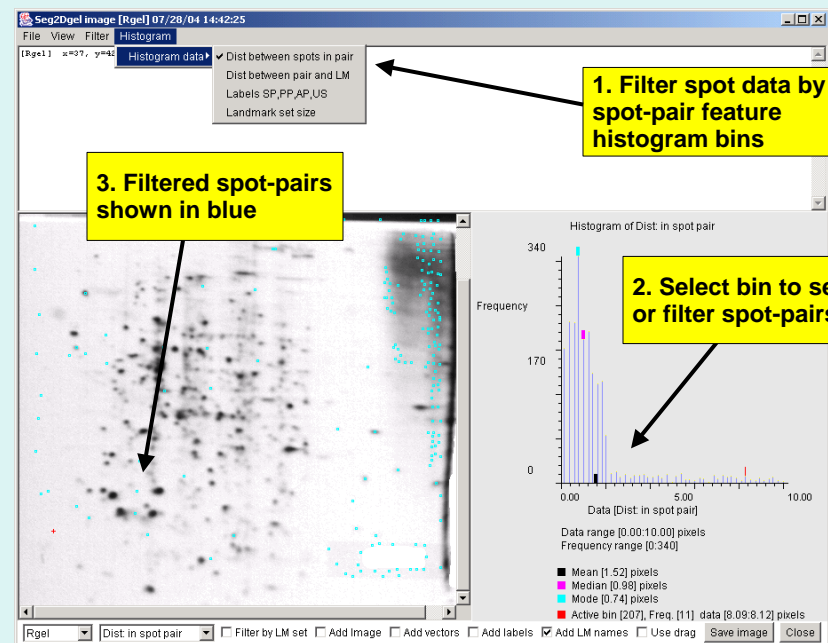
36

Image Viewer - spot-pairs filtered by dP, dist. Between spots

Outlier spot-pairs with histogram dP>8.09 (pixels) in blue

37
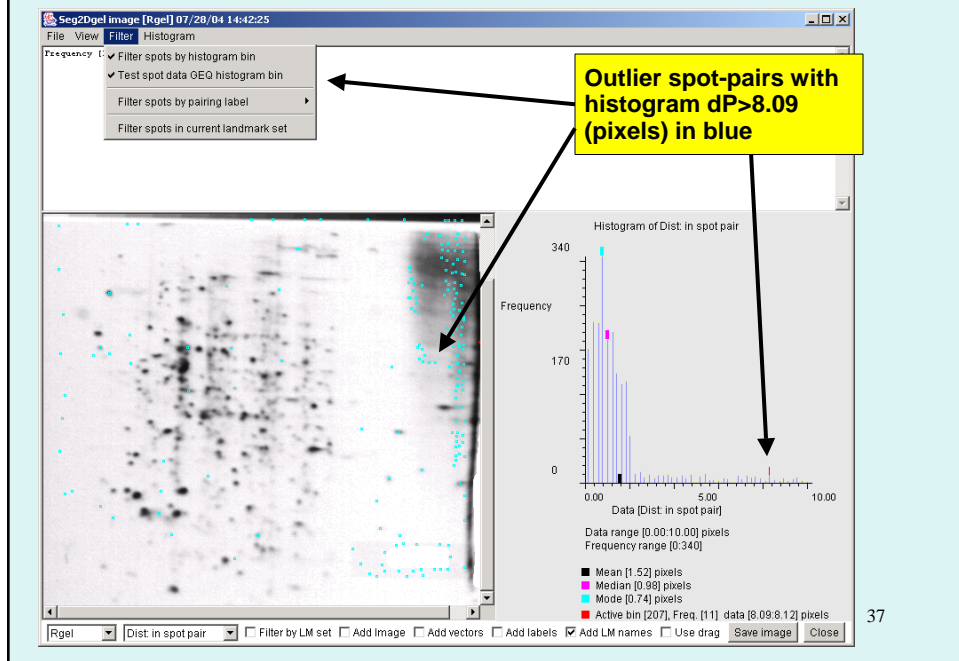
Image Viewer - spot-pairs filtered by dL, dist. LMs to pairs

Outlier spot-pairs with histogram dL>30.32 (pixels) in blue

38
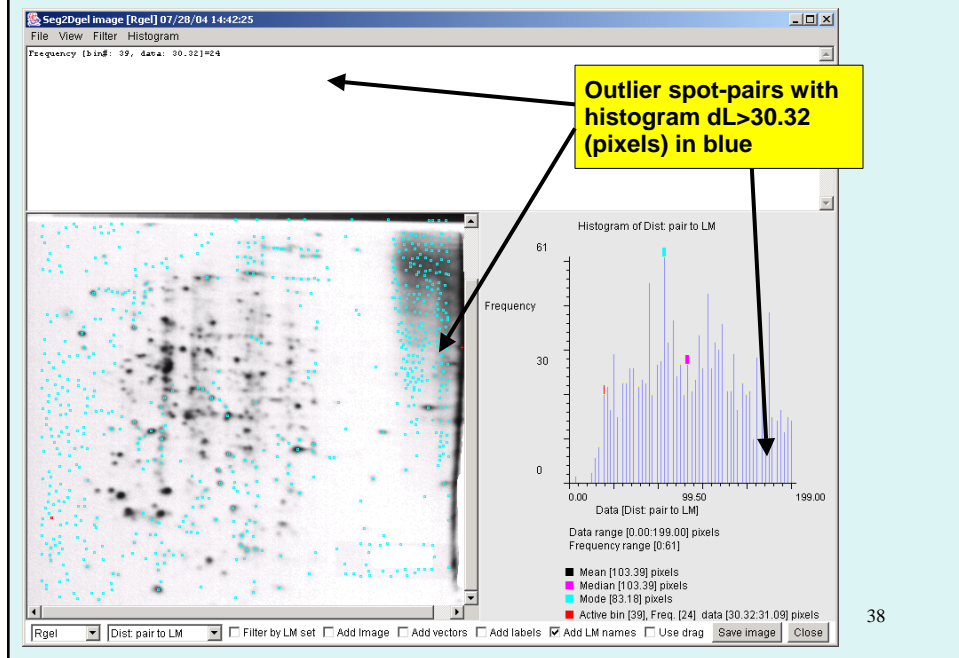
Summary

- CmpSpots is an open-source 2D spot pairing Java program freely available at **http://open2dprot.sourceforge.net/CmpSpots**

- Useful for pairing "spots" in 2D samples (such as 2D PAGE gels, 2D LC-MS and other images with similar types of data).

- It may be used as one of the step [4] alternative modules in the analysis pipeline in the Open2Dprot project at **http://open2dprot.sourceforge.net**

39